

激光与光电子学进展

X 射线荧光光谱结合差分拉曼光谱对烟盒的分类研究

陆润洲¹, 姜红^{1*}, 满吉², 陈娟³, 段斌⁴, 刘峰⁴

¹中国人民公安大学侦查学院, 北京 100038;

²北京华仪宏盛技术有限公司, 北京 100123;

³上海如海光电科技有限公司, 上海 201299;

⁴南京简智仪器设备有限公司, 江苏 南京 210049

摘要 为了建立一种烟盒物证的分类研究方法, 将 X 射线荧光光谱法、拉曼光谱法、差分拉曼光谱法相结合, 对 39 个黄色烟盒样本进行检验。对于 X 射线荧光光谱数据, 采用系统聚类法, 可以将样本初步分为 5 类; 综合拉曼光谱数据与差分拉曼光谱数据, 利用主成分分析法对数据进行降维, 采用 K-Means 聚类在系统聚类的基础上实现更精确的分类, 可以将样本分为 12 类。结果表明, 所提方法充分结合了 X 射线荧光光谱法和差分拉曼光谱法的优势, 结合化学计量学可以对烟盒样本实现较好的分类检验, 为公安机关侦查工作提供线索和方向。

关键词 光谱学; X 射线荧光光谱; 拉曼光谱; 差分拉曼光谱; 烟盒

中图分类号 TS77

文献标志码 A

doi: 10.3788/LOP202259.0530003

Classification of Cigarette Cases by X-Ray Fluorescence Spectrometry Combined with Differential Raman Spectroscopy

Lu Runzhou¹, Jiang Hong^{1*}, Man Ji², Chen Juan³, Duan Bin⁴, Liu Feng⁴

¹College of Investigation, People's Public Security University of China, Beijing 100038, China;

²Beijing Huayi Hongsheng Technology Co., Ltd., Beijing 100123, China;

³Shanghai Oceanhood Opto-Electronics Tech Co., Ltd., Shanghai 201299, China;

⁴Nanjing Jianzhi Instrument and Equipment Co., Ltd., Nanjing, Jiangsu 210049, China

Abstract Thirty-nine yellow cigarette case samples were tested by X-ray fluorescence spectrometry, Raman spectrometry, and differential Raman spectrometry to establish a classification method for cigarette case evidence. Using the hierarchical clustering method, the samples can be classified into five categories based on X-ray fluorescence spectrum data. The Raman spectrum data and differential Raman spectrum data are analyzed by synthesis. The principal component analysis is used to reduce the dimension of the data, followed by K-Means clustering to achieve more accurate classification based on hierarchical clustering. The samples can be classified into 12 categories. The results show that the proposed method fully combines the advantages of X-ray fluorescence spectrometry and differential Raman spectrometry, and combined with chemometrics, it can achieve better classification tests for cigarette case samples and provide clues and direction for the investigation work of public security organs.

Key words spectroscopy; X-ray fluorescence spectrometry; Raman spectrometry; differential Raman spectrometry; cigarette case

收稿日期: 2021-04-07; 修回日期: 2021-05-17; 录用日期: 2021-06-02

基金项目: 中国人民公安大学 2021 年度基科费重点项目(2021JKF212)

通信作者: *jiangh2001@163.com

1 引言

在信息时代,当侦查工作在大数据侦查方向遇到瓶颈时,传统的痕迹物证提取对划定侦查范围、确定侦查方向、锁定犯罪嫌疑人等方面依然具有重要作用^[1]。烟盒是犯罪现场经常出现的痕迹物证之一,在犯罪现场、犯罪现场周围、犯罪嫌疑人路经的地点,都有可能提取到犯罪嫌疑人丢弃的烟盒。但是由于犯罪嫌疑人往往警惕性较高、反侦查意识较强,会在丢弃前对烟盒进行撕碎等处理,并且被丢弃的烟盒的表面往往会受到环境中污秽的污染,因此难以直接从中提取到有用的文字信息,但是烟盒的颜色多数情况下可以识别出来。如果能够对同种颜色的烟盒实现分类鉴别,从而达到对其种属的倾向性认定,便可大大缩小侦查范围,将结果与视频监控排查、购销记录调取等信息化手段相结合,从而为侦查工作提供线索与突破口。

X 射线荧光光谱法通过测试得到烟盒中的元素种类和元素含量,根据元素种类与元素含量的不同实现对烟盒物证的分类鉴别^[2]。拉曼光谱法和差分拉曼光谱法通过测试得到烟盒的拉曼光谱峰位,其拉曼光谱峰位的归属反映了化学键、官能团等分子结构信息,通过比对实现对烟盒物证的分类鉴别^[3]。李春宇等^[4]用 X 射线荧光光谱仪测量 30 种品牌的纸张灰烬,通过测量数据训练支持向量机分类器,最终实现纸张种类和品牌来源的判别。付钧泽等^[5]利用差分拉曼光谱仪对 49 个不同的香烟水松纸样品进行检验,依据填料和着色剂对样品实现初步分类,再借助系统聚类实现对样品的准确分类与区分。郭鹏等^[6]利用拉曼光谱仪,分析 33 个烟用内衬纸样品,根据拉曼光谱图中出峰位置的不同,结合 SPSS 软件实现了对不同品牌烟用内衬纸的区分。这三个实验分别采用了 X 射线荧光光谱法、差分拉曼光谱法和拉曼光谱法这三种不同的测试方法,对测试得到的数据进行分析后都较好地实现了对于样本的分类研究。如果能将不同的光谱方法相结合,可以得到样本更为全面的数据。甄晓宇等^[7]将电感耦合等离子体质谱、红外光谱与 X 射线光谱相结合,测定磁石样品的物质组成、结构及所含元素含量。朱晓晗等^[8]将拉曼光谱法与 X 射线荧光光谱法相结合,实现对鞋套样本的分类研究。X 射线荧光光谱法测试得到的是样本中的元素种类及其含量信息,拉曼光谱法和差分拉曼光谱法反映微观的

分子内部的化学键、官能团等信息,是两种不同的分类依据。如果能够将三种方法相结合,可以实现对样本更为精确的分类。

本文将 X 射线荧光光谱法、拉曼光谱法和差分拉曼光谱法相结合,对 39 个黄色烟盒样本进行测试,比较不同方法之间的区分效果。将 X 射线荧光光谱数据和差分拉曼光谱数据相结合,利用系统聚类、主成分分析法、K-Means 聚类、Fisher 判别分析等手段,实现对烟盒样本的分类研究。

2 实验部分

2.1 实验仪器与实验条件

X-MET8000 Expert 型 X 射线荧光光谱仪, Rh 阳极靶,工作模式为 plastic,电压为 50 kV,电流为 200 μ A,测试时间为 60 s,检测元素范围为 $^{12}\text{Mg}\sim^{92}\text{U}$,工作温度为 $-20\sim 50\text{ }^{\circ}\text{C}$ 。

PR1064 型便携式拉曼光谱仪,上海如海光电科技有限公司,激光光源波长为 1064 nm,激光功率为 500 MW,积分时间设置为 10 s,光谱采集范围为 $100\sim 3000\text{ cm}^{-1}$,电压为 220 V。

SEDRS Portable-Base 型便携式差分拉曼光谱仪,南京简智仪器设备有限公司,选用激发波长分别为 784 nm 和 785 nm 的两种激发光,光谱采集范围为 $180\sim 2800\text{ cm}^{-1}$,激光功率设置为 440 MW,积分时间设置为 10 s。

2.2 实验样本

不同品牌、不同系列的黄色或者是带有部分黄色的烟盒一共 39 种,样本信息如表 1 所示。

实验时截取样本上面 $1\text{ cm}\times 1\text{ cm}$ 大小的黄色部分,用酒精擦拭,晾干。分别用三种仪器在对应的实验条件下对样本进行测试。每个样本平行测定 3 次,取平均值。

2.3 实验数据与谱图的预处理

对于 X 射线荧光光谱仪测试得到的样本中的元素种类和元素含量,通过比对仪器的元素含量检出限,删除属于误差的数据,最终检测出样本中 Cl、Ca、Ti、Fe、Cu、Zn、Cd、Sn、Sb、Hg 这十种元素及对应的元素含量。

对于各样本在 784 nm 和 785 nm 两种波长的激发光下测试得到的差分拉曼光谱数据,利用 back propagation(BP)神经网络技术分离基线偏差和差分信号,结合差分解调和去噪算法,提取出纯净的拉曼光谱^[9]。如图 1 所示,以 6 号样本为例,x 和 y 分

表 1 烟盒样本表

Table 1 Table of cigarette case samples

Label	Brand	Series	Label	Brand	Series
1	Mount Tai	Cigar	21	Mountain Changbai	777
2	Yellow Crane Tower	World famous building	22	Yellow Crane Tower	Hard elegant fragrant gold
3	Lotus King	Hard branchlets	23	Golden Leaf	JinmanTang
4	Good cat	Changle	24	Overseas new Peony	Vietnam
5	Golden Leaf	Golden Eye	25	Diamonds	Good luck
6	Nanjing	95 supreme	26	Lotus	Lotus
7	Nanjing	Riverstones	27	Longyan	Red pine
8	Good cat	Long changle	28	Great Hall of the People	Royal Orchid
9	Nanjing	Jinling twelve hairpins	29	ESSE	Change double
10	Tianzi	Gold	30	Nanjing	Grand View Garden
11	Longyan	Auspicious sign	31	Camel	Star anize hard yellow
12	Mount Huangshan	New concept of Huishang	32	Liqun	Long mouth
13	Yuxi	Sub branch 108	33	Zhonghua	Double middle branches
14	Mount Huangshan	Red square seal	34	Jinsheng	Tengwangge purple light
15	Diamonds	Flos Nelumbinis	35	Mount Hongta	Hard classic 1956
16	Yellow Crane Tower	Hard Canyon	36	Yellow Crane Tower	Hard Canyon tender
17	Mount Huangshan	Middle red square seal	37	Lanzhou	Hard treasures
18	Mount Hongta	Hard classic 100	38	Lanzhou	Fine treasures
19	Yuxi	Hard	39	Marlboro	White golden
20	Golden Leaf	Tianye			

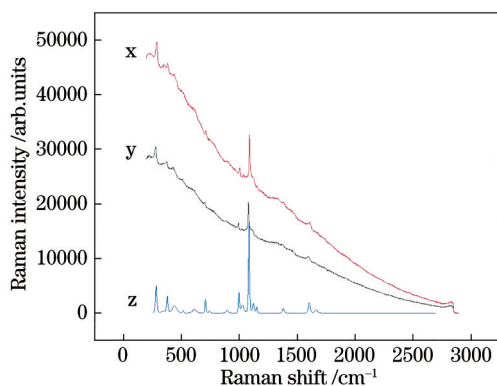


图 1 6号样本的谱图预处理

Fig. 1 Raman spectrum's pretreatment of sample 6

别代表 785 nm 和 784 nm 激发波长条件下测试得到的拉曼光谱, z 代表经过谱图预处理后得到的拉曼光谱。

3 结果与讨论

3.1 系统聚类法

对于 X 射线荧光光谱法测试得到的数据, 在 IBM SPSS Statistics 26.0 软件中, 区间测量选用平方欧氏距离, 聚类方法选用质心法, 对测试结果进行系统聚类, 结果如图 2 所示。由谱系图可知, 当并类距离为 1 时, 样本可分为 9 类; 当并类距离为 2 时, 样品可分为 7 类; 当并类距离为 3 时, 样品可分为

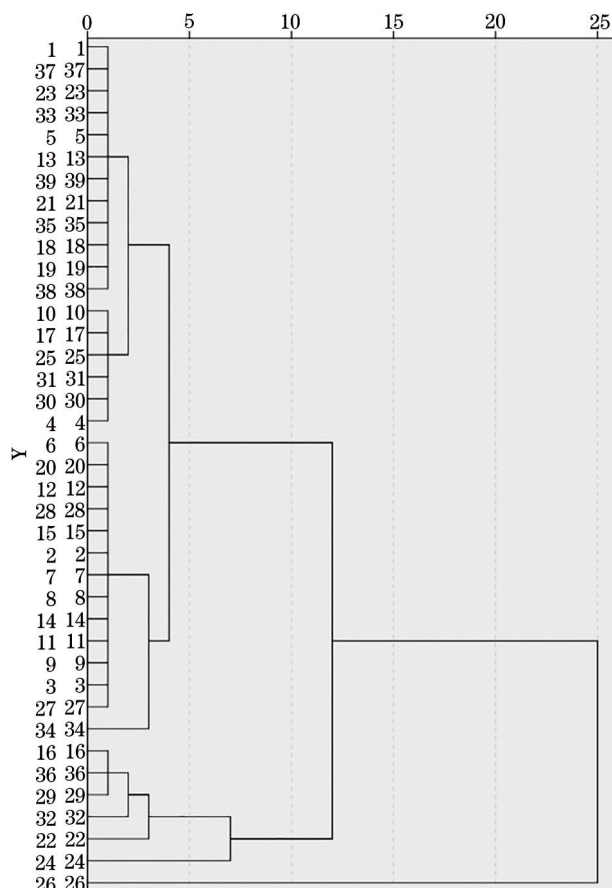


图 2 39种烟盒样本的系统聚类谱系图

Fig. 2 Dendrogram of 39 cigarette case samples using hierarchical clustering

5类。直至阈值达到25时,所有样品并为一类^[10]。

以类别数为横坐标,以聚合系数为纵坐标,绘制折线图,如图3所示,当类别数为5时,折线的下降趋缓,因此通过系统聚类将样本分为5类是最为准确合理的^[11-12]。

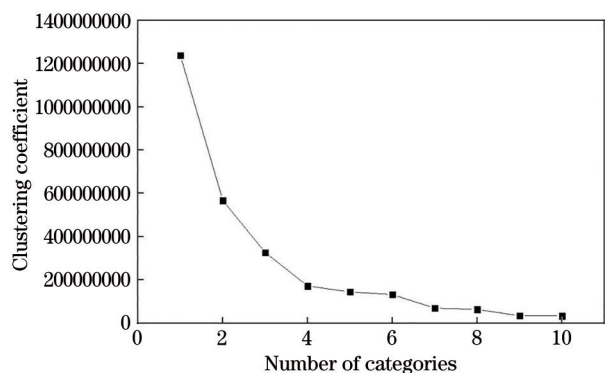


图3 聚合系数折线图

Fig. 3 Line chart of clustering coefficients

系统聚类的结果如表2所示,可以看到,在X射线荧光光谱测试得到的数据的基础上进行系统聚类,分类结果中的I和II两类的样本数量依然较多。

表2 样本系统聚类结果

Table 2 Classification results of samples using hierarchical clustering

Classification	Label
I	1, 4, 5, 10, 13, 17, 18, 19, 21, 23, 25, 30, 31, 33, 35, 37, 38, 39
II	2, 3, 6, 7, 8, 9, 11, 12, 14, 15, 20, 27, 28, 34
III	16, 22, 29, 32, 36
IV	24
V	26

此时,可以借助拉曼光谱与差分拉曼光谱进行分析,以实现对本样本的进一步分类。39个烟盒样本的拉曼光谱如图4所示,可以看到,除了少数几个样本外,样本之间的拉曼光谱区分非常小,在 435 cm^{-1} 、 505 cm^{-1} 、 895 cm^{-1} 、 1333 cm^{-1} 、 1375 cm^{-1} 、 1456 cm^{-1} 这几个峰位处都有明显的特征峰,经查阅资料,这几个峰归属于纤维素的拉曼光谱特征峰^[13],其他的特征峰之间的区分效果较弱,可以根据拉曼光谱将样本分为几大类。而差分拉曼光谱则很好地解决了拉曼光谱信号弱和易受荧光干扰的问题,区分效果更加明显,如图5所示。

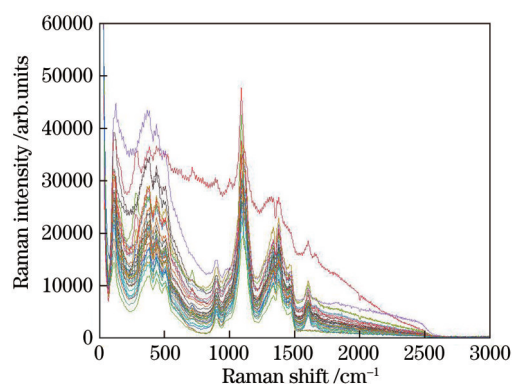


图4 39个烟盒样本的拉曼光谱

Fig. 4 Raman spectra of 39 cigarette case samples

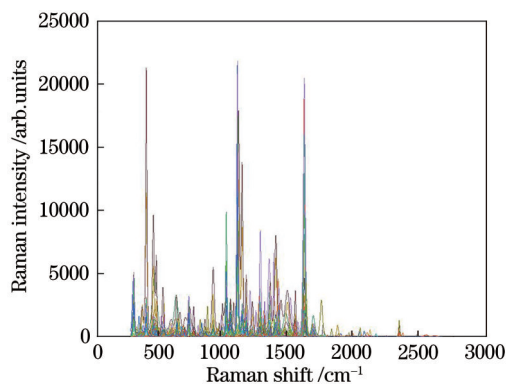


图5 39个烟盒样本的差分拉曼光谱

Fig. 5 Differential Raman spectra of 39 cigarette case samples

3.2 主成分分析法

对测试得到的差分拉曼光谱数据进行分析,以实现对本样本的进一步分类。由于拉曼光谱数据信息量庞大^[14],因此使用主成分分析法对数据进行降维。

设 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)^T$ 是 m 维随机变量,定义均值 $E(\mathbf{X}) = \mathbf{u}$,定义协方差矩阵 $\text{cov}(\mathbf{X}) = \mathbf{\Sigma}$ 。 m 维随机变量 \mathbf{X} 到 m 维随机变量 \mathbf{Y} 的线性变换 $\mathbf{Y}_i = \mathbf{a}_i^T \mathbf{X}$, 其中 $\mathbf{a}_i = (a_{1i}, \dots, a_{mi})^T$, 可以得到 $E(\mathbf{Y}_i) = \mathbf{a}_i^T \mathbf{u}$, $\text{var}(\mathbf{Y}_i) = \mathbf{a}_i^T \mathbf{\Sigma} \mathbf{a}_i$, $\text{cov}(\mathbf{Y}_i, \mathbf{Y}_j) = \mathbf{a}_i^T \mathbf{\Sigma} \mathbf{a}_j$ 。因此,可以定义 $\mathbf{Y}_i = \mathbf{a}_i^T \mathbf{X}$ 为 \mathbf{X} 的第 i 主成分 ($i=1, 2, \dots, m$)。

设 $\mathbf{\Sigma}$ 的特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, 特征值对应的单位特征向量为 $\mathbf{a}_1, \dots, \mathbf{a}_m$, 则 \mathbf{X} 的第 k 个主成分为 $\mathbf{Y}_k = \mathbf{a}_k^T \mathbf{X}$, 方差为 $\text{var}(\mathbf{Y}_k) = \mathbf{a}_k^T \mathbf{\Sigma} \mathbf{a}_k$ 。求解一个主成分等价于求解一个最优化问题,以第一主成分为例:

$$\max_{\mathbf{a}_1} \mathbf{a}_1^T \mathbf{\Sigma} \mathbf{a}_1, \quad (1)$$

式中: $\mathbf{a}_1^T \mathbf{a}_1 = 1$ 。定义拉格朗日函数:

$$\mathbf{a}_1^T \boldsymbol{\Sigma} \mathbf{a}_1 - \lambda (\mathbf{a}_1^T \mathbf{a}_1 - 1), \quad (2)$$

式中： λ 为拉格朗日乘子。将拉格朗日函数对 \mathbf{a}_1 求导，并令其为 $\mathbf{0}$ ，得

$$\boldsymbol{\Sigma} \mathbf{a}_1 - \lambda \mathbf{a}_1 = \mathbf{0}. \quad (3)$$

因此， λ 是 $\boldsymbol{\Sigma}$ 的特征值， \mathbf{a}_1 是对应的单位特征向量，因此有

$$\mathbf{a}_1^T \boldsymbol{\Sigma} \mathbf{a}_1 = \mathbf{a}_1^T \lambda \mathbf{a}_1 = \lambda \mathbf{a}_1^T \mathbf{a}_1 = \lambda \mathbf{I}. \quad (4)$$

假设 \mathbf{a}_1 是 $\boldsymbol{\Sigma}$ 的最大特征值 λ_1 对应的单位特征向量，则此问题的最优化解为 \mathbf{a}_1 和 λ_1 ，所以第一主成分为 $\mathbf{a}_1^T \mathbf{X}$ ，其方差为协方差矩阵的最大特征值：

$$\text{var}(\mathbf{a}_1^T \mathbf{X}) = \mathbf{a}_1^T \boldsymbol{\Sigma} \mathbf{a}_1 = \lambda_1 \mathbf{I}. \quad (5)$$

在本实验中，样本数量为39，拉曼光谱数据有2408维，则观测数据的样本矩阵为 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{39})^T \in \mathbf{R}^{39 \times 2408}$ ，样本的协方差矩阵为 $\mathbf{S} =$

$\sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T / (39 - 1)$ ，其中 $\bar{\mathbf{x}} = 39^{-1} \mathbf{x}$ ，样本的相关矩阵为 $\text{diag}(\mathbf{S})^{-1/2} \mathbf{S} \text{diag}(\mathbf{S})^{-1/2}$ 。则样本的第一主成分 $\mathbf{a}_1^T \mathbf{x}$ 就是在 $\mathbf{a}_1^T \mathbf{a}_1 = 1$ 的条件下，使 $\mathbf{a}_1^T \mathbf{x}_i (i = 1, 2, \dots, 39)$ 的样本方差 $\mathbf{a}_1^T \mathbf{S} \mathbf{a}_1$ 最大的 \mathbf{x} 的线性变换。同理，可求出样本其他主成分。

选取主成分数量的常用准则有崖底碎石图、累计方差贡献率和Kaiser准则^[15]。在累计方差贡献率的标准中，累计方差贡献率一般要求大于85%。每个主成分的累计方差贡献率为 $\lambda_i / \sum_{i=1}^m \lambda_i$ ，因此前 k 个主成分的累计方差贡献率为 $\sum_{i=1}^k \lambda_i / \sum_{i=1}^m \lambda_i$ 。通过数据处理最终提取到14个主成分，且累计贡献率为90.575%，如表3所示，即通过主成分分析法将2408维数据降低到14维。降维后的数据将用于后续的处理^[16-17]。

表3 总方差解释

Table 3 Total variance interpretation

Component	Initial eigenvalue			Extraction sum of squared loadings		
	Total	Variance / %	Cumulative value / %	Total	Variance / %	Cumulative value / %
1	610.234	26.954	26.954	610.234	26.954	26.954
2	290.041	12.811	39.765	290.041	12.811	39.765
⋮	⋮	⋮	⋮	⋮	⋮	⋮
13	44.159	1.95	88.921	44.159	1.95	88.921
14	37.45	1.654	90.575	37.45	1.654	90.575

3.3 K-Means 聚类

为了实现对 I、II、III 三类样本更为精确的分类，采用K-Means法对这三类样本分别进行进一步的分类。对于K-Means法中 K 值的选取，即聚类数的选取，采用轮廓系数法^[18]。某个样本点 X_i 的轮廓系数 S 定义为

$$S = \frac{b - a}{\max(a, b)}, \quad (6)$$

式中： a 是凝聚度，是 X_i 与同簇的其他样本的平均距离； b 是分离度，是 X_i 与最近簇中所有样本的平均距离。最近簇是指，用 X_i 到某个簇所有样本平均距离作为衡量该点到该簇的距离，离 X_i 最近的一个簇。最近簇 C_j 表示为

$$C_j = \arg \min_{C_k} \frac{1}{n} \sum_{p \in C_k} |p - X_i|^2, \quad (7)$$

式中： p 是某个簇 C_k 中的样本。

对所有样本的轮廓系数取平均值，得到平均轮廓系数。簇内样本的距离越近，簇间样本距离越远，平均轮廓系数越大，聚类效果越好。因此，平均

轮廓系数最大的 K 便是最佳聚类数。

对 I 和 II 两类样本分别进行轮廓系数的计算，得到最佳的 K 值分别为5和4。对其进行K-Means聚类，最终得到的结果如表4所示。

分析聚类结果可以发现，Golden Leaf牌5号和23号两个样本都在I-2类，Lanzhou牌37号和38号两个样本都在I-3类，Nanjing牌6号、7号和9号三个样本都在II-3类，Yellow Crane Tower牌16号、22号和36号三个样本都在III类。I-3类样本多为细支香烟的包装，I-5类样本多为粗支香烟包装，II-1类样本多为细支烤烟型的包装。可以看出，所提方法对样本类别的倾向性认定效果非常好。

3.4 Fisher 判别分析

可以在烟盒样本的分类结果基础上建立预测模型，以实现对新样本类别变量取值的预测。常见的预测方法有Logistic回归模型^[19]、CART决策树模型^[20]、Fisher判别分析^[21]等。Fisher判别法是在方差分析的基础上建立的一种区分各个总体的线性判别法，将多维数据投影到低维空间，实现数据

表 4 使用 K-Means 聚类对样本进行分类的结果
Table 4 Classification results of samples using K-Means clustering

Classification	Label
I-1	18
I-2	5, 23
I-3	1, 13, 17, 19, 30, 31, 33, 35, 37, 38
I-4	10
I-5	4, 21, 25, 39
II-1	2, 3, 8, 20, 34
II-2	14, 15, 27
II-3	6, 7, 9, 11, 12
II-4	28
III	16, 22, 29, 32, 36
IV	24
V	26

类内距离的尽可能聚拢,类间距离的尽可能分离^[22]。以 X 射线荧光光谱数据为基础得到的分类结果和以差分拉曼光谱数据为基础得到的分类结果都可以进行 Fisher 判别分析,以前者为例,表 5 是 Fisher 判别函数的非标准化的系数矩阵,4 个判别函数分别为 $Y_1 = -0.001X_1 + 0.04X_5 - 0.016X_6 + 0.026X_7 + 0.018X_8 - 6.993$, $Y_2 = 0.001X_1 + 0.001X_2 + 0.001X_3 - 0.009X_5 - 0.011X_6 + 0.002X_7 - 0.017X_8 - 11.214$, $Y_3 = -0.001X_2 - 0.001X_3 - 0.004X_4 + 0.056X_5 + 0.044X_6 - 0.007X_7 - 0.019X_8 - 5.315$, $Y_4 = 0.001X_2 + 0.003X_3 + 0.001X_4 - 0.054X_5 - 0.054X_6 - 0.001X_7 - 0.03X_8 + 3.815$ 。其中,变量 X_1 为 Cl 元素含量, X_2 为 Fe 元素含量, X_3 为 Cu 元素含量, X_4 为 Zn 元素含量, X_5 为 Cd 元素含量, X_6 为 Sn 元素含量, X_7 为 Sb 元素含量, X_8 为 Hg 元素含量。

表 5 典型判别函数系数
Table 5 Canonical discriminant function coefficients

Element	Y_1	Y_2	Y_3	Y_4
Cl	-0.001	0.001	0.000	0.000
Ca	0.000	0.000	0.000	0.000
Ti	0.000	0.000	0.000	0.000
Fe	0.000	0.001	-0.001	0.001
Cu	0.000	0.001	-0.001	0.003
Zn	0.000	0.000	-0.004	0.001
Cd	0.040	-0.009	0.056	-0.054
Sn	-0.016	-0.011	0.044	-0.054
Sb	0.026	0.002	-0.007	-0.001
Hg	0.018	-0.017	-0.019	-0.030
Constant	-6.993	-11.214	-5.315	3.815

但是,这四种判别函数并不是每一种都能很好地实现对类别变量取值的预测,如表 6 所示,第一判别函数的特征值为 12.566,累计百分比为 52.3%,第二判别函数的特征值为 10.883,累计百分比为 97.6%,其余两个判别函数的解释方差能力较低,仅在 0 至 2.4% 之间。结果说明,第一判别函数和第二判别函数明显优于其他两个判别函数。

表 6 判别函数特征值
Table 6 Eigenvalues of discriminant functions

Function	Eigenvalue	Variance / %	Cumulative value / %	Canonical correlation
Y_1	12.566 ^a	52.3	52.3	0.962
Y_2	10.883 ^a	45.3	97.6	0.957
Y_3	0.428 ^a	1.8	99.4	0.548
Y_4	0.150 ^a	0.6	100	0.361

Note: a indicates that the first four typical discriminant functions are used in the analysis.

为了验证判别函数一和判别函数二预测效果更好的结论,可以采用威尔克的 λ (Wilks' λ) 统计量反向测度。

$$\text{Wilks}'\lambda_t = \prod_{i=t}^I \frac{1}{1 + E_i}, \quad (8)$$

式中: t 为第 t 个判别函数; I 为最后一个判别函数; E 为特征值。可以得出, I 值越小,整体判别能力越强。

检验统计量是在威尔克的 λ 基础上构建的:

$$X_t^2 = - \left(n - \frac{P + G}{2} - 1 \right) \ln \text{Wilks}'\lambda_t, \quad (9)$$

式中: X_t 为检验统计量; n 为总的样本量; P 为判别变量个数; G 为类别数; $\ln \text{Wilks}'\lambda_t$ 小于 0。在本实验中, n 为 39, P 为 8, G 为 8。该统计量近似服从卡方分布。当概率值小于显著性水平 0.05 时,认为当前判别函数整体的判别能力较强。表 7 中, Chi-square、df、Sig. 的物理含义分别为卡方、自由度、显著性,它们用于检测不同判别函数判别未知样本的效果,即检测判别函数区分样本的显著程度。函数检验 1 直至 4, 检验统计量观测值为 170.161, 对应的概率值为 0; 函数检验 2 直至 4, 检验统计量观测值为 90.63, 对应的概率值为 0。剩余的概率值都大于 0.05, 因此可以考虑舍弃第三和第四判别函数^[23-25]。

所以,选择函数一和函数二作为判别轴建立联合分布图,如图 6 所示。在联合分布图中,每个点的横坐标比纵坐标的区分效果更加明显,即判别函数一比判别函数二的预测效果更好。因此,在本实验中,判别函数一可以实现对黄色烟盒样本的判别分

表 7 威尔克的 λ
Table 7 Wilks' λ

Test of functions	Wilks' λ	Chi-square	df	Sig.
1 through 4	0.004	170.161	40	0
2 through 4	0.051	90.63	27	0
3 through 4	0.609	15.14	16	0.514
4	0.869	4.267	7	0.749

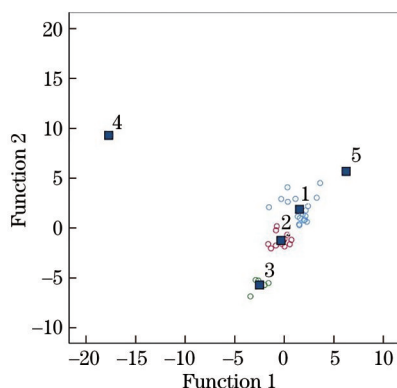


图 6 Fisher 判别函数分布图
Fig. 6 Fisher discriminant function distribution

类,同理,同样可以对差分拉曼光谱数据进行判别分析,实现对新样本类别变量取值的预测。

4 结 论

对于 39 个不同品牌、不同系列的黄色烟盒样本,分别用 X 射线荧光光谱法、拉曼光谱法和差分拉曼光谱法对其进行测试。首先根据 X 射线荧光光谱测试得到的元素种类和元素含量,结合系统聚类法,可以将样本分为 5 类。综合分析拉曼光谱数据与差分拉曼光谱数据,利用主成分分析法对差分拉曼光谱数据进行降维,对于系统聚类结果中样本数量较多的类别,依据降维后的差分拉曼光谱数据,结合 K-Means 聚类法,最终可以将样本分为 12 类。结合样本实际情况分析,分类的结果对于样本类别的倾向性认定效果非常好。采用 Fisher 判别分析,在样本分类的基础上建立预测模型,可以实现对样本类别倾向性取值的预测。

对于以后在犯罪现场及周边提取到的烟盒物证,可以按照本文的思路,对烟盒物证的类别的倾向性认定提供帮助,从而为案件侦破寻找突破口。根据其所属的类别种类,大大缩小侦查范围,结合大数据手段进行综合分析,排查可能的销售地点,结合现场视频监控调取等手段,为犯罪嫌疑人的追踪、案件的侦破提供帮助。

参 考 文 献

- [1] Zhou Z. The importance of traditional evidence of material evidence by investigating a burglary case[J]. Legality Vision, 2016(26): 123.
周舟. 由一起入室盗窃案的侦破谈传统痕迹物证的重要性[J]. 法制博览, 2016(26): 123.
- [2] Chen Z H, Li J Q, Zhao Z L, et al. Establishment and application of testing method for fluorescence based soft X-ray absorption spectrum[J]. Acta Optica Sinica, 2019, 39(3): 0330002.
陈振华, 李俊琴, 赵子龙, 等. 软 X 射线荧光吸收谱测试方法的建立与应用[J]. 光学学报, 2019, 39(3): 0330002.
- [3] Lü Q H, Wang X H, Shen A G, et al. Application of Raman spectroscopy techniques in on-site fast detection[J]. Journal of Instrumental Analysis, 2019, 38(5): 612-617.
吕前辉, 王小华, 沈爱国, 等. 拉曼光谱技术在现场快检分析领域中的应用[J]. 分析测试学报, 2019, 38(5): 612-617.
- [4] Li C Y, Liu J K, Jiang H, et al. Identification of X-ray fluorescent spectral paper ashes based on support vector machine algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(3): 032801.
李春宇, 刘金坤, 姜红, 等. 基于支持向量机算法的 X 射线荧光光谱纸张灰烬识别研究[J]. 激光与光电子学进展, 2021, 58(3): 032801.
- [5] Fu J Z, Jiang H, Liu F, et al. Discrimination of cigarette tipping paper by differential Raman spectroscopy combined with system clustering[J]. Chemical Research and Application, 2020, 32(11): 1973-1978.
付钧泽, 姜红, 刘峰, 等. 差分拉曼光谱结合系统聚类检验香烟水松纸[J]. 化学研究与应用, 2020, 32(11): 1973-1978.
- [6] Guo P, Jiang H. A study on analysis of inner liner for cigarette by Raman[J]. China Pulp & Paper Industry, 2016, 37(12): 53-58.
郭鹏, 姜红. 拉曼光谱检验烟用内衬纸的研究[J]. 中华纸业, 2016, 37(12): 53-58.
- [7] Zhen X Y, Zheng C, Pi Z F, et al. Study on a quality control method for Mongolian medicine magnetite based on inductively coupled plasma mass spectrometry combined with infrared spectroscopy and X-ray diffraction[J]. Journal of Instrumental Analysis, 2021, 40(1): 112-118.
甄晓宇, 郑重, 皮子凤, 等. 电感耦合等离子体质谱

- 结合红外光谱与 X 射线光谱的蒙药磁石质量控制方法研究[J]. 分析测试学报, 2021, 40(1): 112-118.
- [8] Zhu X H, Jiang H. X-ray fluorescence spectrometry combined with Raman spectroscopy analysis for disposable plastic shoe covers[J]. Chemical Research and Application, 2019, 31(11): 1927-1930.
朱晓晗, 姜红. 拉曼光谱法结合 X 射线荧光光谱法检验一次性塑料鞋套的研究[J]. 化学研究与应用, 2019, 31(11): 1927-1930.
- [9] Fang G, Yin L, Liu F, et al. Application research of fluorescence suppression based on differential Raman technique[J]. Laser Technology, 2019, 43(3): 359-362.
方刚, 殷磊, 刘峰, 等. 基于差分喇曼技术在抑制荧光中的应用研究[J]. 激光技术, 2019, 43(3): 359-362.
- [10] Ma X, Jiang H, Yang J Q. Examination of plastic pack belts(ropes) via X-ray fluorescence spectrometry combined with multivariate statistical analysis[J]. Laser & Optoelectronics Progress, 2019, 56(22): 223005.
马泉, 姜红, 杨佳琦. X 射线荧光光谱结合多元统计分析塑料打包带(绳)[J]. 激光与光电子学进展, 2019, 56(22): 223005.
- [11] Zhang N N, Sun S J, Zhang Y J. Application of unascertained clustering in patent quality evaluation [J]. Modern Electronics Technique, 2020, 43(8): 143-146.
张妮妮, 孙胜娟, 张永健. 未确知聚类在专利质量评价中的应用[J]. 现代电子技术, 2020, 43(8): 143-146.
- [12] Zhang N N. Comprehensive evaluation of patent quality based on unascertained clustering[D]. Handan: Hebei University of Engineering, 2020.
张妮妮. 基于未确知聚类的专利质量综合评价[D]. 邯郸: 河北工程大学, 2020.
- [13] Ma J F, Yang S M, Tian G L, et al. Study on the application of Raman spectroscopy to the research on natural cellulose structure[J]. Spectroscopy and Spectral Analysis, 2016, 36(6): 1734-1739.
马建锋, 杨淑敏, 田根林, 等. 拉曼光谱在天然纤维素结构研究中的应用进展[J]. 光谱学与光谱分析, 2016, 36(6): 1734-1739.
- [14] Cao X Y, Jin S Z, Hou B, et al. Pollen detection and classification method via Raman spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(13): 133001.
曹馨艺, 金尚忠, 侯彬, 等. 基于拉曼光谱的花粉检测及分类方法[J]. 激光与光电子学进展, 2020, 57(13): 133001.
- [15] Yin B. A new method to determine the number of principal component extraction [J]. Statistics & Decision, 2010(19): 8-10.
尹波. 主成分抽取数量确定的新方法[J]. 统计与决策, 2010(19): 8-10.
- [16] Zhu X H, Hu Y N, He X Y, et al. Detection of disposable paper cup by Raman spectroscopy and PCA[J]. China Pulp & Paper, 2020, 39(9): 38-42.
朱晓晗, 胡越宁, 何歆沂, 等. 基于主成分分析法对一次性纸杯的拉曼光谱检验研究[J]. 中国造纸, 2020, 39(9): 38-42.
- [17] Song H S, Ma L Z, Wang Y F, et al. Recognition of formaldehyde, methanol based on PCA-BP neural network[J]. Laser & Optoelectronics Progress, 2020, 57(7): 071201.
宋海声, 麻林召, 王一帆, 等. 基于 PCA-BP 神经网络对甲醛和甲醇的识别研究[J]. 激光与光电子学进展, 2020, 57(7): 071201.
- [18] Wang X H. An automatic approach to solving clustering problems with the number of clusters unknown based on the particle swarm optimization and silhouette coefficient[J]. Journal of Yunnan Minzu University (Natural Sciences Edition), 2016, 25(4): 367-371.
王学贺. 一种基于改进微粒群和轮廓系数的划分聚类方法[J]. 云南民族大学学报(自然科学版), 2016, 25(4): 367-371.
- [19] Jin X, Zhu X Z, Li S W, et al. Predicting soil available phosphorus by hyperspectral regression method based on gradient boosting decision tree[J]. Laser & Optoelectronics Progress, 2019, 56(13): 131102.
金秀, 朱先志, 李绍稳, 等. 基于梯度提升树的土壤速效磷高光谱回归预测方法[J]. 激光与光电子学进展, 2019, 56(13): 131102.
- [20] Iwabuchi Y, Kameyama M, Matsusaka Y, et al. A diagnostic strategy for Parkinsonian syndromes using quantitative indices of DAT SPECT and MIBG scintigraphy: an investigation using the classification and regression tree analysis[J]. European Journal of Nuclear Medicine and Molecular Imaging, 2021, 48(6): 1833-1841.
- [21] He Y, Wang J F. Rapid nondestructive identification of wood lacquer using Raman spectroscopy based on characteristic-band-Fisher-K nearest neighbor[J]. Laser & Optoelectronics Progress, 2020, 57(1): 013001.
何亚, 王继芬. 基于特征波段-Fisher-K 近邻的木器漆拉曼光谱的快速无损鉴别[J]. 激光与光电子学进展, 2020, 57(1): 013001.

- [22] Li X R, Hu Z Y, Zhao Y H. Supervised feature extraction based on FDA and galaxy spectra classification[J]. Spectroscopy and Spectral Analysis, 2007, 27(9): 1898-1901.
李乡儒, 胡占义, 赵永恒. 基于 Fisher 判别分析的有监督特征提取和星系光谱分类[J]. 光谱学与光谱分析, 2007, 27(9): 1898-1901.
- [23] Yan W J, Wei C J, Fan L Y, et al. Study on microscopic laser Raman spectroscopy of automobile lampshade based on Fisher discriminant-support vector machine[J]. China Plastics, 2021, 35(3): 124-129.
颜文杰, 卫辰洁, 范琳媛, 等. 基于 Fisher 判别-支持向量机的汽车灯罩显微激光拉曼光谱研究[J]. 中国塑料, 2021, 35(3): 124-129.
- [24] Fu J Z, Jiang H, Li Y, et al. Examination of cigarette ash evidence by XRF combined with chemometrics[J]. Laser & Optoelectronics Progress, 2021, 58(6): 063003.
付钧泽, 姜红, 李意, 等. XRF 结合化学计量学检验香烟烟灰物证[J]. 激光与光电子学进展, 2021, 58(6): 063003.
- [25] Zhu X H, Jiang H, Cui A S, et al. Spectrum identification of disposable plastic gloves based on Fisher discriminant analysis[J]. China Plastics Industry, 2020, 48(7): 108-112, 116.
朱晓晗, 姜红, 崔傲松, 等. 基于 Fisher 判别分析的一次性塑料手套光谱鉴别[J]. 塑料工业, 2020, 48(7): 108-112, 116.